European Journal of Operational Research 255 (2016) 110–120

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Stochastics and Statistics

# A multi-kernel support tensor machine for classification with multitype multiway data and an application to cross-selling recommendations

Zhen-Yu Chen[a,*], Zhi-Ping Fan[a,b], Minghe Sun[c]

[a] Department of Information Management and Decision Sciences, School of Business Administration, Northeastern University, Shenyang 110167, China
[b] State Key Laboratory of Synthetical Automation for Process Industries, Northeastern University, Shenyang 110819, China
[c] Department of Management Science and Statistics, College of Business, The University of Texas at San Antonio, San Antonio, TX 78249-0632, USA

ARTICLE INFO

*Article history:*
Received 17 July 2015
Accepted 10 May 2016
Available online 17 May 2016

*Keywords:*
Data mining
Customer relationship management
Multitype multiway data
Support tensor machine
Cross-selling

ABSTRACT

Cross-selling is an integral component of customer relationship management. Using relevant information to improve customer response rate is a challenging task in cross-selling recommendations. Incorporating multitype multiway customer behavioral, including related product, similar customer and historical promotion, data into cross-selling models is helpful in improving the classification performance. Customer behavioral data can be represented by multiple high-order tensors. Most existing supervised tensor learning methods cannot directly deal with heterogeneous and sparse multiway data in cross-selling recommendations. In this study, a novel collaborative ensemble learning method, multi-kernel support tensor machine (MK-STM), is proposed for classification in cross-selling recommendations using multitype multiway customer behavioral data. The MK-STM can also perform feature selections from large sparse multitype multiway data. Computational experiments are conducted using two databases. The experimental results show that the MK-STM has better performance than existing ensemble learning, supervised tensor learning and other commonly used recommendation methods for cross-selling recommendations.


## 1. Introduction

Cross-selling has become an integral component of customer development in the life cycle of customer relationship management (CRM) (Kamakura, 2008; Ngai, Xiu, & Chau, 2009). Cross-selling refers to promotion activities aiming at selling products to customers who have already bought some other products from the same vendor (Knott, Hayes, & Neslin, 2002; Li, Sun, & Montgomery, 2011; Ngai et al., 2009). Selling additional products to the same customers can help the firm increase the customer lifetime value, improve the relationship with customers and reduce the chance of churn (Kamakura, 2008; Prinzie & Van den Poel, 2006). The development of electronic commerce and the availability of massive customer data make cross-selling play a more important role in the future business activities. Cross-selling has been a widely adopted practice in electronic commerce, *e.g.*, product bundling recommendation at Amazon and service packaging by online travel companies (Netessine, Savin, & Xiao, 2006). Cross-selling managers face challenging tasks of improving the low customer response rate and avoiding "over-touching" the customers (Kamakura, 2008; Li et al., 2011). Consequently, it is very important in cross-selling recommendations to offer the right product to the right customer at the right time. Cross-selling recommendation is usually treated as a rating prediction, classification, ranking, or top-K recommendation problem (Adomavicius & Tuzhilin, 2005; Adomavicius & Zhang, 2016; Kamakura, 2008). It is treated as a classification problem in this study.

Customer demographic and behavioral sequential data are usually used in cross-selling modeling (Kamakura, 2008). Unlike repeated purchases, there are only historical behavioral data of related products (services) and similar customers available for cross-selling recommendations. Introducing and using diverse customer behavioral data from multiple views may help in improving the recommendation performance and the customer response rate. In general, three types of customer behavioral data can be used for cross-selling recommendations as listed in the following.

(1) Related product data. Customer purchase history of related products, *i.e.*, products that are similar or complementary to the target product, can be used to predict the purchase likelihood of the target product. Related product data have been used in classification models for cross-selling

* Corresponding author. Tel.: +86 24 83656354.
*E-mail addresses:* zychen@mail.neu.edu.cn (Z.-Y. Chen), zpfan@mail.neu.edu.cn (Z.-P. Fan), minghe.sun@utsa.edu (M. Sun).

http://dx.doi.org/10.1016/j.ejor.2016.05.020
0377-2217/

recommendations (Kamakura, Kossar, & Wedel, 2004; Prinzie & Van den Poel, 2006, 2007, 2011). Moreover, similarities between products have been used in item-based collaborative filtering (I-CF) methods to make recommendations (Adomavicius & Tuzhilin, 2005).

(2) Similar customer data. Purchase history of similar customers, *i.e*., customers with similar purchase behavior to the target customer, can be used to predict the purchase likelihood of the target customer. Similarities between customers have been used in user-based collaborative filtering (U-CF) methods to make recommendations (Adomavicius & Tuzhilin, 2005).

(3) Historical promotion data. Past promotions may have long-term effects on customer purchase behavior (Li et al., 2011). A few studies have used historical promotion data as variables in cross-selling models (Li et al., 2011).

These three types of data can be treated as longitudinal behavioral data describing certain behavior of customers at certain times. Prinzie and Van den Poel (2006, 2007, 2011) described customer purchase behavior as unidimensional or multivariate sequences without considering the time aspect and used sequence analysis techniques to predict the next product that a customer may purchase. Unlike the sequential data, longitudinal behavioral data (Chen, Fan, & Sun, 2012) have fixed time-intervals, and thus maybe used to predict both the likelihood and the timing of the next purchase of specific products for cross-selling. Moreover, cross-selling recommendations may not result in the immediate sales (Li et al., 2011). Thus, historical promotion data are also necessary to be treated as longitudinal behavioral data. Three typical purchase behavioral, *i.e*., recency, frequency and monetary (RFM), variables can be extracted for each time-interval and used as the derived variables for these three types of data.

The related product, similar customer and historical promotion data have higher dimensions than the customer demographic and aggregated behavioral data. They are of multiway data and are represented by high-order tensors (Hoff, 2011; Sun, Tao, Papadimitriou, Yu, & Faloutsos, 2008). A tensor that generalizes the notions of vectors (first-order tensors) and matrices (second-order tensors) is a natural way of representing multiway data (Signoretto, De Lathauwer, & Suykens, 2011; Sun et al., 2008). The related product, similar customer and historical promotion data are represented as a fourth-order tensor, a fifth-order tensor and a fourth-order tensor, respectively, in this study. Because multiple types of multiway data are used, the term “multitype multiway data” is used to represent the input data of the ensemble learning models in this study.

Compared with multiway data in many other applications such as image and medical signal processing, multiway data in cross-selling recommendations have two distinct characteristics, *i.e*., heterogeneousness and sparseness. Business firms usually record and store large amount of heterogeneous customer data in their data warehouses (Chen, Fan, & Sun, 2012, 2015a). As mentioned above, the classification models in cross-selling recommendations involve three types of multiway data with different orders. To the best of our knowledge, no supervised tensor learning methods can be directly applied to multitype multiway data. Moreover, the uses of large amount of longitudinal behavioral data in cross-selling recommendations provide both opportunities to improve the classification performance and challenges to deal with redundant data. Hence, it is important to develop sparse tensor learning methods to identify potential sparse structures of multitype multiway data. Most existing supervised tensor learning methods cannot filter multiway data and end up with using sparse representations.

In the last decade, support vector machine (SVM) (Chen, Fan, & Sun, 2012; Lessmann & Voß, 2009; Vapnik, 1995) and multiple kernel learning (MKL) (Bach, Lanckriet, & Jordan, 2004; Chen, Fan, & Sun, 2012; Gönen & Alpaydın, 2011) have been hot research topics in machine learning and data mining and have been successfully applied to many fields. Specifically, multi-kernel SVM (MK-SVM) is a state-of-the-art ensemble learning method which combines multiple heterogeneous data and improves classification performance (Chen, Fan, & Sun, 2012, 2015a). Moreover, MK-SVM has good scalability to incorporate new types of data.

A novel MKL method, multi-kernel support tensor machine (MK-STM), is proposed for cross-selling recommendations using multitype multiway data in this study. Unlike other supervised tensor learning methods, the MK-STM can directly deal with multitype multiway data. Furthermore, the MK-STM, as a selective ensemble method, can select a subset of features, *i.e*., variables, with good discriminative abilities from a large number of variables in the sparse multitype multiway data. Moreover, a collaborative ensemble learning framework is developed to apply the ensemble learning methods for classification with multitype multiway data.

This article is organized as follows. Section 2 discusses the relevant literature and outlines the contributions of this study. Section 3 describes the tensor representation of the input data and the collaborative ensemble learning framework. The MK-STM for cross-selling recommendations using multitype multiway data is developed in Section 4. The computational experiments are described and computational results are reported in Section 5. Conclusions and directions for further research are given in Section 6. The preliminaries of multilinear algebra and three typical supervised tensor learning methods are described in the online Appendices.

## 2. Relevant literature

This study is related to two fields of research in the literature, *i.e*., cross-selling and supervised tensor classification. These two fields will be discussed briefly and the contributions of this study relative to these fields will be outlined.

Unlike other elements of CRM and direct marketing such as customer segmentation, customer targeting and churn management, there are relatively few studies on cross-selling (Kumar, George, & Pancras, 2008; Ngai et al., 2009; Prinzie & Van den Poel, 2006). As discussed by Kamakura (2008), the analytical methods for cross-selling can be grouped into acquisition pattern analysis and collaborative filtering (CF). For acquisition pattern analysis, the data of previous purchases of the current and other related customers are used to identify the next product to recommend (Kamakura, 2008). Kamakura et al. (2004) developed a multivariate split-hazard model for multidimensional acquisition pattern analysis to estimate the probability and timing of purchasing new products. Prinzie and Van den Poel (2006, 2007, 2011) considered customer purchase behaviors as unidimensional or multivariate sequences, and respectively applied the mixture transition distribution model, the Markov chain and the Bayesian network to model the behavioral data and predict the next purchase of a customer. Ansell, Harrison, and Archibald (2007) combined the customer lifestyle segmentation and the proportional hazard model to identify cross-selling opportunities using the demographics and the first five purchases of a customer. Kumar et al. (2008) considered aggregated behavioral characteristics, marketing effort and product characteristics in a statistical model to identify the drivers of cross-selling and to target customers. Ahn, Ahn, Oh, and Kim (2011) used demographic and aggregated behavioral data as input of multiple classification models for cross-selling in the mobile telecom market and used genetic algorithms to find solutions. For CF, associations of purchases across customers and items are used to identify cross-selling opportunities and to recommend additional products to customers (Bellogin, Cantador, & Castells, 2013; Kamakura, 2008). In the